Forecasting stock prices with long-short term memory neural network based on attention mechanism

A Mini-project Report submitted in partial fulfillment of the requirements for the award of the degree of

Master of Science in Computer Science

by

Sayan Roy

(19419CMP019)

Department of Computer Science

Institute of Science

Banaras Hindu University, Varanasi – 221005

February 2021

CANDIDATE’S DECLARATION

I **Sayan Roy** hereby certify that the work, which is being presented in the Mini-project report, entitled **Forecasting stock prices with long-short term memory neural network based attention mechanism**, in partial fulfillment of the requirement for the award of the Degree of **Master of Science in Computer Science** and submitted to the institution is an authentic record of my/our own work carried out during the period Nov-2020 to Feb-2021 under the supervision of **Prof. Suresh Selvam**. I also cited the reference about the text(s) /figure(s) /table(s) /equation(s) from where they have been taken.

The matter presented in this Mini-project as not been submitted elsewhere for the award of any other degree or diploma from any Institutions.

Date: Signature of the Candidate

The Viva-Voce examination of *Sayan Roy*, M.Sc. (Computer Science) Student has been held on ________________.

Signature of the Supervisor

ABSTRACT

Stock market is the most complex and volatile market, everyone always try to predict the accurate and efficient way to trade stock. The main goal of this work is the implementation of long-short term memory (LSTM) neural network using Recurrent Neural Network (RNN). Long-short term memory (LSTM) neural network using recurrent neural network (RNN) is used in various fields. LSTM avoids long-term issues due to its unique storage and unit structure, it helps to predict financial time series.

Here we are using HSI (Hang Seng Index) where data are from 2nd January 2002 to 1st July 2019. Data are indexed in the timed order.

Keywords: Open, High, Low, Close, Adj. Close, and Volume.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure No. **Title** **Page No.**

LIST OF ABBREVIATIONS

OPEN	The price of an individual at which a stock started trading.
HIGH	Highest price at which stock traded during a period.
CLOSE	Price of an individual stock when the stock exchange closed.
DATE	in which date the stock has been traded in the exchange.
LOW	Lowest price at which stock traded during a period.
VOLUME	Data Encryption Standard
ADJ. CLOSE	closing price to reflect that it's value after accounting for any corporate actions. It is also used for examining historical returns or for a detailed analysis of past performance.

CHAPTER 1
INTRODUCTION

1.1 GENERAL

Stock market is a public market for trading of company stock. It allows us to buy or sell stocks. A stock maybe bought or sold once it is enlisted in a stock exchange (Exchange is a place where stockbrokers and traders can buy and sell stocks). Compared to other investment product (Bonds and fixed deposits) stock investing provide investors an excellent possibility of making greater return amount in a comparatively shorter period of time. As the economy grows so does the corporate earnings. Because of economic growth, the average income grows.

1.2 Objectives

As there is a lot of chances of gaining some extra profit as well as there is a chances to lose the money in the market. It is time consuming to invest in a stock, need to research a lot about that stock and there is many formalities in the process. Investing in stock is subjected to many risk, market is too volatile. The stocks of a company goes up and comes down so many times in just a single day. These price fluctuations are unpredictable most of the times and investor sometimes have to face severe loss due to such uncertainty.

1.3 RNN

Recurrent Neural Network (RNN) are a type of Neural Network where the output from previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in case like when it is required to predict the next price of a stock price , the previous prices are required and hence there is a need to remember the previous price. Thus RNN came in play, which is able to solve this issue with the help of hidden layer. The important feature of RNN is Hidden State, Which remembers some information about a particular sequence.

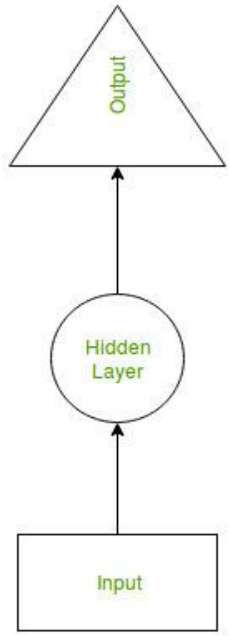

Fig-.1: RNN (Recurrent neural network)

RNN have a memory which remembers all the information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity unlike other neural networks.

Table 1.1: Comparison report of Advantages and Disadvantages.

Technique	Advantages	Disadvantages
	• An RNN remembers each and every information through time. It is useful in time series prediction only because of the feature to remember previous input as well. This is called Long Short term Memory.	• Gradient Vanishing and exploding problems.
	• Recurrent neural network are even used with convolutional layers to extend the effective pixel neighbourhood.	• Training an RNN is a very difficult task.
		• It cannot process very long sequences if using tanh or relu as an activation function.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

LSTM uses one of the most common forms of RNN. This time recurrent neural network meant to avoid long-term dependence problems and is suitable for processing and predicting time series. Proposed by Sepp Hochreiter and Jurgen Schmidhuber in 1997, the LSTM model consists of a unique set of memory cells that replace the hidden layer neurons of the RNN, and its key is the state of the memory cells. The LSTM model filters information through the gate structure to maintain and update the state of memory cells.

2.2 LSTM

The forgotten gate in the LSTM unit determines which cell state information is discarded from the model. According to the figure-2, the memory cell accepts the output h_{t-1} of the previous moment and external information X_t of the current moment as inputs and combines then into a long vector $[h_{t-1}, X_t]$ through σ transformation to become

$$f_t = \sigma\ (W_f\ * [h_{t-1}, x_t] + b_f), \qquad (1)$$

Where $w_{f\ is}$ and b_f are the weight matrix and bias of the forgotten gate and σ is the sigmoid function. The main function of forgotten gate is to record how much cell state Ct-1 of the previous time is reserved to the cell state Ct of the current time. The output of the gate will be between 0 and 1 based on the value of ht-1 and xt where 1 indicates complete reservation and 0 indicate the complete discernment.

The input gate determines how much of the current time in network input xt is reserved into the cell state Cp which prevents content from entering into the memory cells. Basically It has two functions one is to find the state of the cell that must be updated and the value to need to be updated in the selected sigmoid later, as equation-2 and the other thing is to update the information in the cell state. A new Candidate vector Ct is created through the tanh layer to control the new information which will be added as equation -3. And the equation 4 is for updating the cell state of the memory cells.

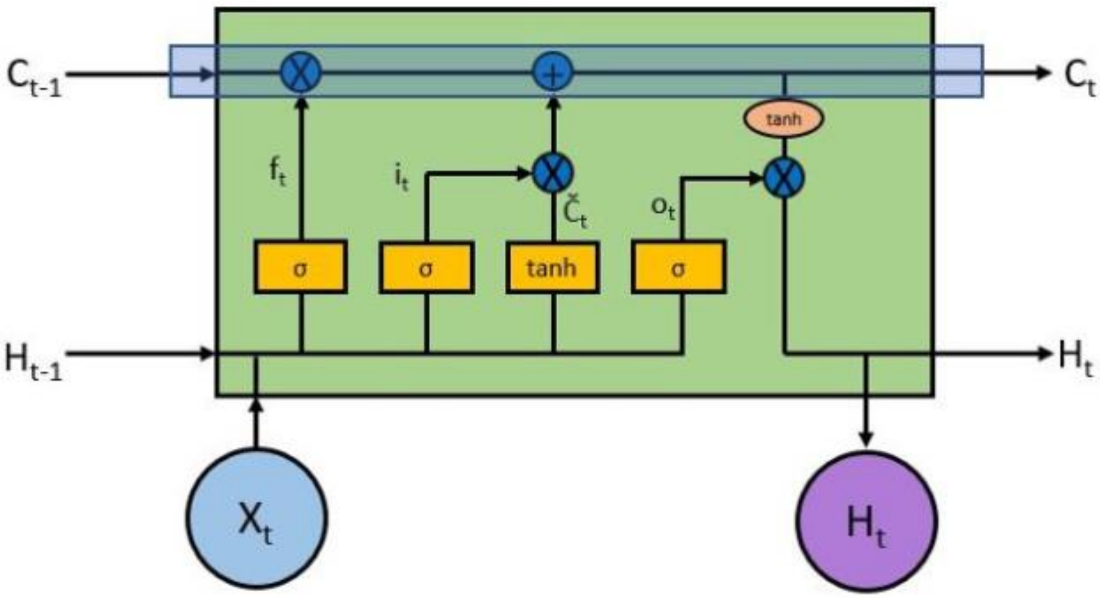

Fig-2 Structure of long short term memory (LSTM)

$$i_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_i), \tag{2}$$

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t. \tag{4}$$

$$O_t = \sigma(W_\sigma \cdot [h_{t-1}, x_t] + b_o). \tag{5}$$

$$h_t = O_t * \tanh(C_t). \tag{6}$$

2.3 Attention Mechanism

Many algorithms and mechanisms are inspired by biological phenomena. For example inspired by astrocytes in the biological nervous system that can greatly regulate the operation of neurons, song et al. proposed a spiking neural P system with gel-like control functions. Like that derived from the study of human vision the attention mechanism highlights the most important local information by allocating an adequate attention to key information. The attention mechanism is an excellent in the case of serialized data, speech recognition, machine translation and part-of-speech tagging. It is also used with RNN to classify images focusing on the important parts to reduce the task's complexity. It has been also applied to machine translation to enable the simultaneous translation and alignment.

The attention mechanism is applied in the stock forecasting through the extraction of the information in the news in an auxiliary role to compare the price fluctuations. Liu proposed an attention mechanism to based cyclic neural network to train the financial news to predict the stock prices. It can have either soft or hard attention. The hard attention mechanism aims on one element in the input information, selecting such information based on either maximum or random sampling which requires more training to obtain good results. Whereas the soft attention mechanism assigns weight to all the input information to enables more efficient use of input information and to obtain results in timely manner. This soft attention mechanism can be like as.

$$e_t = \tanh(w_a[x_1, x_2, \ldots, x_T] + b) \tag{7}$$

$$a_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)}, \tag{8}$$

Where wa is the weight of the attention mechanism, which indicates the information that should be emphasized, et is the result of the first weighing calculations, b is the deviation of the attention mechanism.[x1,x2…..xt] is the input for the attention mechanism. At is then final weight obtained by [x1,x2,….xt]

We calculate the degree of matching of each element in the input information and the input the matching degree to the softmax function to generate the attention distribution. The general mechanism has two steps: 1) calculate the attention distribution, 2) calculate the weight of the average of the input information according to the attention distribution and finally attention weight vector is weighted and averaged with the input information to obtain the final result. The attention value is obtained as shown in fig-3.

2.4 Wavelet transform

Wavelet analysis has remarkable achievements in the field of image and signal processing. Its ability to compensate for the shortcomings of Fourier analysis, day by day it has been introduced in the economic and financial fields. The wavelet transform has unique advantages in solving the traditional time series and analysis problems. It can decompose and reconstruct financial time series data from different time and frequency domain scales. Therefore, combining both of them has enable us to better analyze and resolve problem in financial time series.

Financial time series can be consider as a signal. Wavelet threshold denoising has the basic idea to wavelet transform a signal, where the wavelet coefficient of the noise generated by wavelet decomposition is smaller than that of the signal. A threshold value is to be selected to separate the useful signal from the noise and then the noise is set to zero. The basics steps to wavelet decomposition, threshold processing and reconstruction of signals.

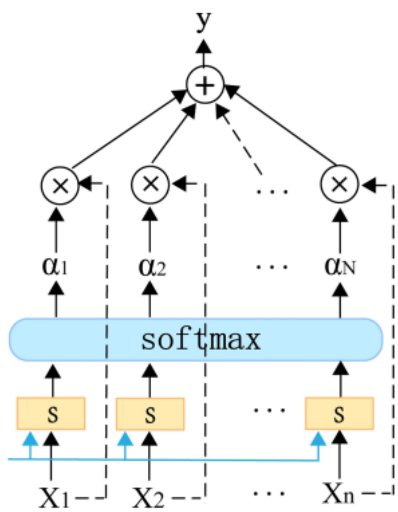

Fig-3 Basic Structure of Attention model.

It depends on four factors: 1) selection of wavelet basis function; 2) determination of the number of decomposition layers; 3) determination of the threshold value; and 4) selection of the threshold function. Commonly used wavelet basis functions which are also suitable for financial data denoising are Haar, db, N, Sym N.

CHAPTER 3

PROPOSED APPROACH

3.1 Introduction

To implement a stock index price forecasting model has three stages to follow: data collection and preprocessing, model establishment and training and evaluation of the experimental results as shown in the fig-4. Fig-5 the LSTM-attention network structure consists of data input, hidden layers and output layers and hidden layer consists of an LSTM, attention and dense layer.

3.1.1 Data source

We have selected stock indices as experimental data on Hang Seng index (HSI). The HSI data are from January 2, 2002 to July 1, 2019.

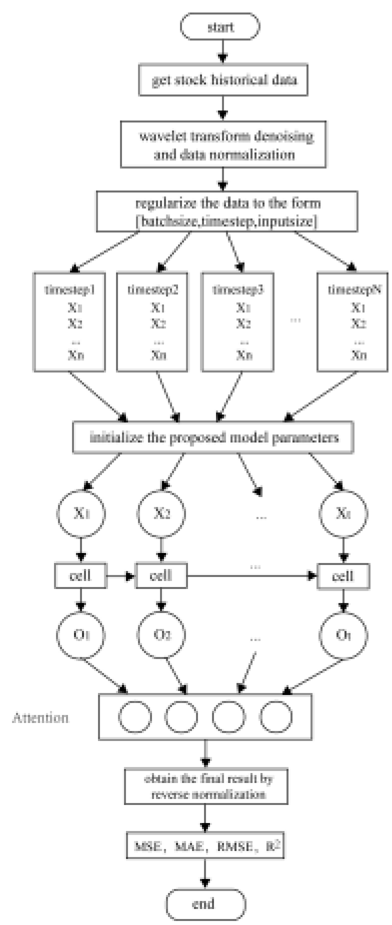

Fig-4 Attention-based LSTM model flowchart

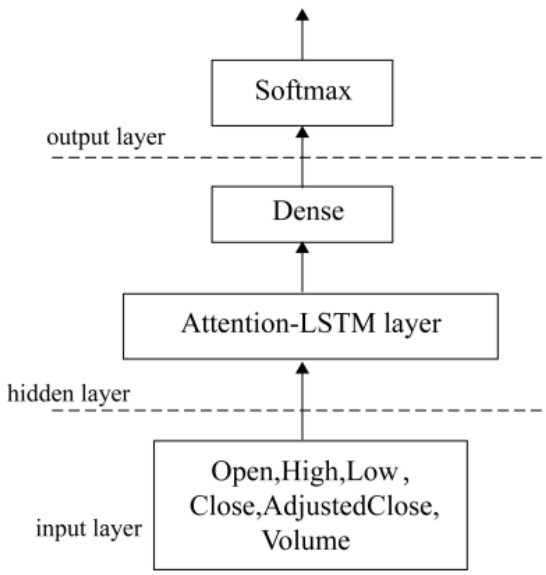

Fig-5 LSTM Attention network structure.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2002/1/2	11368.12988	11368.12988	11241.62988	11350.84961	11350.84961	106074600
1	2002/1/3	11348.79981	11447.00000	11348.79981	11423.51953	11423.51953	274626200
2	2002/1/4	11547.00977	11727.80957	11547.00977	11702.15039	11702.15039	362553600
3	2002/1/7	11687.42969	11905.54981	11687.42969	11892.63965	11892.63965	442387000
4	2002/1/8	11781.19043	11806.25000	11678.05957	11713.70996	11713.70996	279504600

Table2 Partial Stock Data Samples for HIS (Hang Seng Index)

The data are from yahoo finance. Total six basic variables in the basic dataset. The opening price (Open) is the first transaction per share after the market open and stock trades start. (Close) Closing price it’s the final price of that day. High is the highest price a stock trades in a day and low is the lowest price in a day. Adjusted close is the closing price after adjustments for splits and dividend distributions. Data are adjusted using appropriate split and dividend multipliers. Volume is the number of transactions in a time unit for a transaction. We have listed some data samples in the table-2

Chapter-4

IMPLEMENTATION

4.1 Data preprocessing

We have implemented the proposed stock forecasting method using python TensorFlow. We have used zero-mean normalization to the data and divided it into training and test datasets. For the HSI datasets. Which is used to training the model and also use for testing.
Due to the complex and volatile stock market and various trading restrictions, the stock prices we see are noisy. At the same time, the financial time series is nonstationary and exhibits the overlapping of useful signals and noise, which makes traditional denoising ineffective. The wavelet transformation is considered more suitable for extremely irregular financial sequences because it can perform both time domain and frequency domain analysis. It combines with the traditional theory of time series analysis and shows good applicability. Therefore, wavelet transform tool becomes a powerful tool to process financial time series data. We use a wavelet transform with a multi-scale characteristic to denoise the dataset and separating the useful signal from the noise. More specifically we use the coif3 wavelet function with three decomposition layers, and we evaluate the effect of the wavelet transform by its signal-to-noise (SNR) and root mean square error (RMSE). As the higher SNR the smaller will be the RMSE,

$$\mathrm{SNR} = 10\log\left[\frac{\sum_{j=1}^{N} x_j^{\,2}}{\sum_{j=1}^{N} (x_j - \hat{x}_j)^2}\right]. \tag{9}$$

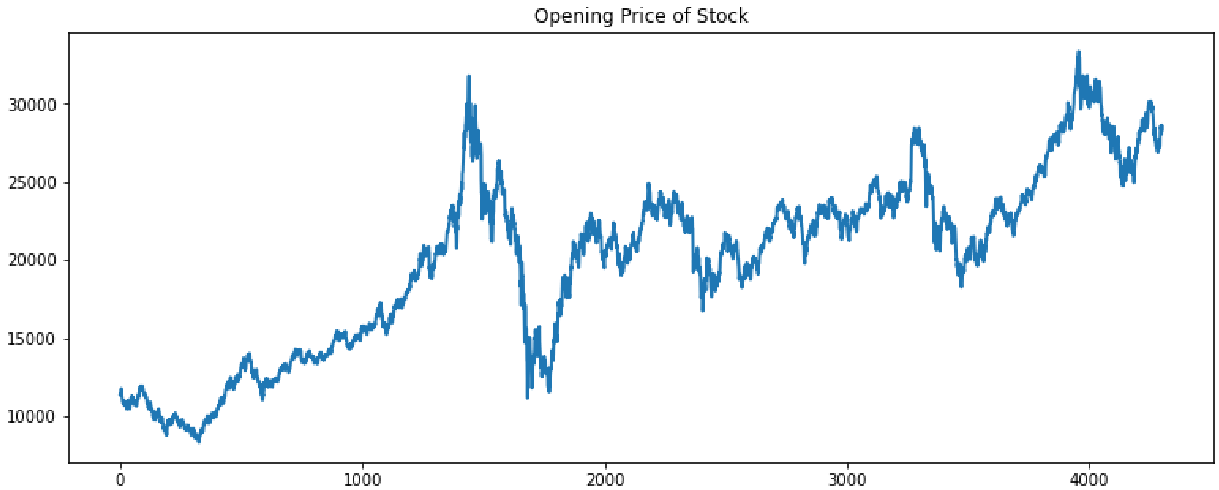

Fig-6 Opening price without denoising

4.2 LSTM-Attention Model

Input data include two types of data, stock price and volume, we use standardization and normalization to process the data and to improve the training effect of the neural network model. Finally, the denoising data are the input to the LSTM-Attention training model. The data are normalized to the form [B, T, D], where B is the batch size, T is the time step, and D is the dimension of the input data. The representation will be like this,

$$\begin{bmatrix} x_1^{(1)} & \cdots & x_1^{(D)} \\ \vdots & \ddots & \vdots \\ x_T^{(1)} & \cdots & x_T^{(D)} \end{bmatrix}.$$

Each matrix is used to represent input data for a time stamp. The parameters of the model are initialized and the process input data are sequentially transmitted to the cells in the LSTM layer.

Take the output from the previous cell and use it as input to the attention layer. The proposed model network structure is an LSTM cyclic network with 100 hidden nodes per layer. The learning rate is set to 0.0001 and the number of iterations is 100.

4.3 Model performance metrics

We have evaluated the prediction results and the established prediction model by the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2). The smaller the MSE, RMSE, and MAE, the closer the predicted value to the true value; the closer the coefficient R^2 to 1, the better the fit of the model

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\hat{\mathrm{y}}_\mathrm{i} - \mathrm{y}_\mathrm{i}), \tag{10}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} |(\hat{\mathrm{y}}_\mathrm{i} - \mathrm{y}_\mathrm{i})|, \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\hat{\mathrm{y}}_\mathrm{i}} - \mathrm{y}_\mathrm{i}), \tag{12}$$

Fig-7 Opening price after denoising

Chapter-5

5.1 Results and discussions

We have experimented with different wavelet functions and used SNR and RMSE values to determine which wavelet was more suitable for stock price denoising. From the result we have found out that the SNR values of coif3 were the largest and the RMSE values were the smallest among the four wavelet functions. Therefore we choose coif3 as the wavelet function for the experiment. Fig-6 shows the opening price curve before denoising using the wavelet transform. Fig-7 shows the opening price curve after denoising using the wavelet function. By comparing the two, it is found that the noise after wavelet transform processing is smaller.

We processed stock index data datasets in the LSTM model: the LSTM (WLSTM) model with the wavelet transform, the gate recurrent unit (GRU) neural network model, and our prosed WLSTM+Attention model. We trained them and compare the predicted results.

However, on the HIS dataset, although the proposed model is superior to the others, the error and model fit are significantly worse than on the other two datasets. Different datasets may make the model have different performance. As can be seen in table 3 and 4, the model performs better on the U.S stock forecast.

Fig-8 Forecast result of LSTM on HSI data (green is the predicted data, orange is test data and blue is input data)

Chapter-6

6.1 Conclusion & Future work:

From the paper we can establishes a forecasting framework to predict the opening prices of stocks. We process the stock data through a wavelet transform and used an attention-based LSTM neural network to predict the stock opening price, with excellent results.

This work found that an attention based LSTM has more predictive outcomes for price prediction than other methods. However, only considering the historical data on price trends is too singular and many not be able to fully accurately forecast the price on a given day. Therefore, we can add data predictions related to stock-related news and basic information, so to enhance the stability and accuracy of the model in the case of major event.

REFERENCES

[1] Zhhi S U, Man L U, Dexuan L I. Deep Learning in Financial Empirical Applications: Dynamics, Contributions and Prospects [J]. Journal of Financial Research, 2017.

[2] Bin Weng, Ahmed M A , Megahed F M. Stock Market One-day ahead Movement Prediction using Disparate Data Sources[J]. Exper Systems with Applications, 2017, 79(2):153-163.

[3] Xu T, Zhang J, Ma Z, et al. Deep LSTM for Large Vocabulary Continuous Speech Recognition [J]. 2017.

[4] Kim J, El-Khamy M, Lee J. Residual LSTM: Design of a Deep Recurrent Architecture for Distant Speech Recognition [J]. 2017

[5] Shih C H, Yan B C, Liu S H, et al. Investigating Siamese LSTM networks for text categorization[C]//Asia-pacific Signal & Information Processing Association Summit & Conference. 2017.

[6] Simistira F, Ul-Hasan A, Papavassiliou V, et al. Recognition of Historical Greek Polytonic Scripts Using LSTM Networks[C]// 13th International Conference on Document Analysis and Recognition. 2015.

[7] Bontempi G, Taieb S B, Yann-Ae¨l Le Borgne. Machine learning strategies for time series forecasting[C]// European Business Intelligence Summer School. ULB—Universite Libre de Bruxelles, 2013.

[8] Tay F E H, Cao L. Application of support vector machines in financial time series forecasting [J]. Journal of University of Electronic Science & Technology of China, 2007, 29(4):309–317.

[9] Lin Y, Guo H, Hu J. An SVM-based approach for stock market trend prediction[C]// The 2013 International Joint Conference on Neural Networks (IJCNN). IEEE, 2013.

[10] Wanjawa B W, Muchemi L. ANN Model to Predict Stock Prices at Stock Exchange Markets [J]. Papers, 2014.

[11] Zhang R, Yuan Z, Shao X. A New Combined CNN-RNN Model for Sector Stock Price Analysis[C]//2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). IEEE Computer Society, 2018.

[12] Ha Young Kim,Chang Hyun Won. Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models [J]. Expert Systems With Applications, 2018,103.

[13] Liu S, Zhang C, Ma J. CNN-LSTM Neural Network Model for Quantitative Strategy Analysis in Stock Markets[C]// International Conference on Neural Information Processing. Springer, Cham, 2017.

[14] Zhao Z, Rao R, Tu S, et al. Time-Weighted LSTM Model with Redefined Labeling for Stock Trend Prediction[C]// 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE,2017.

[15] Jiang Q, Tang C, Chen C, et al. Stock Price Forecast Based on LSTM Neural Network[J]. 2018.

[16] Jin Z, Yang Y, Liu Y. Stock Closing Price Prediction Based on Sentiment Analysis and LSTM[J]. Neural Computing and Applications, 2019(3).

[17] Lecun Y, Bengio Y, Hinton G. Deep learning.[J]. Nature, 2015, 521(7553):436. https://doi.org/10.1038/nature14539 PMID: 26017442

[18] Abedinia O, Amjady N, Zareipour H.A new feature selection technique for load and price forecast of electrical power systems[J].IEEE Transactions on Power Systems, 2017, 32 (1):62–74.

[19] Song T, Pan L. Spiking neural P systems with request rules[J]. Neurocomputing, 2016: S0925231216002034, Pages 193–200.

[20] Song T, Gong F, Liu X, et al. Spiking Neural P Systems with White Hole Neurons[J]. IEEE Transactions on NanoBioscience, 2016, 15(7):1–1.

[21] Song T, Zheng P, Wong M L D, et al. Design of Logic Gates Using Spiking Neural P Systems with Homogeneous Neurons and Astrocytes-like Control[J]. Information Sciences, 2016, 372:380–391.

[22] Mnih V, Heess N, Graves A, et al. Recurrent Models of Visual Attention[J]. Advances in neural information processing systems, 2014.

[23] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.

[24] Liu Huicheng. Leveraging Financial News for Stock Trend Prediction with Attention-Based RecurrentNeural Network[J]. 2018.

[25] Wang Q, Li Y, Liu X. Analysis of Feature Fatigue EEG Signals Based on Wavelet Entropy[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2018, 32(8):1854023.

[26] Sun Y, Shen X, Lv Y, et al. Recaptured Image Forensics Algorithm Based on Multi-Resolution Wavelet Transformation and Noise Analysis[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2017, 32(02):1790–1793.

[27] Ramsey JB. The contribution of wavelets to the analysis of economic and financial data. Philosophical Transactions of the Royal Society B Biological Sciences. 1999; 357(357):2593–606.

[28] Dessaint O, Foucault T, Fre´sard Laurent, et al. Noisy Stock Prices and Corporate Investment[J]. Social Science Electronic Publishing;2018.

[29] Chang S, Gupta R, Miller S M, et al. Growth Volatility and Inequality in the U.S.: A Wavelet Analysis[J].Working Papers, 2018.

[30] Ardila D, Sornette D. Dating the financial cycle with uncertainty estimates: a wavelet

proposition[J].Finance Research Letters, 2016:S1544612316301593.